

PartA

November 29, 2024

```
[7]: import csv
import re
import pandas as pd
df = pd.read_csv('nba.csv')
df = df.fillna('Unknown')
print(df)
df.to_csv('nba_modified.csv', index=False)
df1 = pd.read_csv('nba1.csv')
df1 = df1.fillna('Unknown')
print(df1)
df1.to_csv('nba1_modified.csv', index=False)
def validate_data(row):
    column_rules = {
        "Unnamed": lambda x: re.match(r'^-?\d+$', x) or x == "Unknown",
        "height": lambda x: re.match(r'^-?\d+$', x) or x == "Unknown",
        "weight": lambda x: re.match(r'^-?\d+$', x) or x == "Unknown",
        "born": lambda x: re.match(r'^-?\d+$', x) or x == "Unknown",
        "Player": lambda x: x.isalpha(),
        "collage": lambda x: x.isalpha(),
        "birth_city": lambda x: x.isalpha(),
        "birth_state": lambda x: x.isalpha(),
    }
    for column, rule in column_rules.items():
        if column in row:
            value = row[column].strip()
            if rule(value) is False:
                return False
    return True
with open('Players(processed).csv', 'r') as file:
    reader = csv.DictReader(file)
    is_valid = True
    for i, row in enumerate(reader):
        if i <= 10000:
            continue
        if not validate_data(row):
            is_valid = False
            print(f"Invalid data found in row {i}: {row}")
```

```

        break
if is_valid:
    print("All data is valid.")

```

	Rk	Player	Age	Team	Pos	G \
0	1.0	Joel Embiid	28.0	PHI	C	66.0
1	2.0	Luka Dončić	23.0	DAL	PG	66.0
2	3.0	Damian Lillard	32.0	POR	PG	58.0
3	4.0	Shai Gilgeous-Alexander	24.0	OKC	PG	68.0
4	5.0	Giannis Antetokounmpo	28.0	MIL	PF	63.0
..
675	536.0	Deonte Burton	29.0	SAC	PF	2.0
676	537.0	Michael Foster Jr.	20.0	PHI	PF	1.0
677	538.0	Frank Jackson	24.0	UTA	PG	1.0
678	539.0	Alondes Williams	23.0	BRK	SG	1.0
679	Unknown	League Average	Unknown	Unknown	Unknown	Unknown

	GS	MP	FG	FGA	...	DRB	TRB	AST \
0	66.0	34.6	11.0	20.1	...	8.4	10.2	4.2
1	66.0	36.2	10.9	22.0	...	7.8	8.6	8.0
2	58.0	36.3	9.6	20.7	...	4.0	4.8	7.3
3	68.0	35.5	10.4	20.3	...	4.0	4.8	5.5
4	63.0	32.1	11.2	20.3	...	9.6	11.8	5.7
..
675	0.0	3.0	0.0	1.0	...	0.0	0.0	0.0
676	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0
677	0.0	5.0	0.0	3.0	...	1.0	2.0	1.0
678	0.0	5.0	0.0	0.0	...	1.0	1.0	0.0
679	Unknown	Unknown	Unknown	Unknown	...	Unknown	Unknown	Unknown

	STL	BLK	TOV	PF	PTS	Awards \
0	1.0	1.7	3.4	3.1	33.1	MVP-1DPOY-9CPOY-5ASNBA1
1	1.4	0.5	3.6	2.5	32.4	MVP-8CPOY-8ASNBA1
2	0.9	0.3	3.3	1.9	32.2	CPOY-10ASNBA3
3	1.6	1.0	2.8	2.8	31.4	MVP-5CPOY-7ASNBA1
4	0.8	0.8	3.9	3.1	31.1	MVP-3DPOY-6ASNBA1
..
675	0.0	0.0	0.0	0.0	0.0	Unknown
676	0.0	0.0	0.0	0.0	0.0	Unknown
677	0.0	0.0	0.0	0.0	0.0	Unknown
678	0.0	0.0	2.0	1.0	0.0	Unknown
679	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

Player-additional

0	embiijo01
1	doncilu01
2	lillada01
3	gilgesh01

```

4          antetgi01
..          ...
675        burtode02
676        fostemi02
677        jacksfr01
678        willlial06
679        -9999

```

[680 rows x 32 columns]

	Rk	Player	Age	Team	Pos	G \
0	1.0	Joel Embiid	28.0	PHI	C	66.0
1	2.0	Luka Dončić	23.0	DAL	PG	66.0
2	3.0	Damian Lillard	32.0	POR	PG	58.0
3	4.0	Shai Gilgeous-Alexander	24.0	OKC	PG	68.0
4	5.0	Giannis Antetokounmpo	28.0	MIL	PF	63.0
..
675	536.0	Deonte Burton	29.0	SAC	PF	2.0
676	537.0	Michael Foster Jr.	20.0	PHI	PF	1.0
677	538.0	Frank Jackson	24.0	UTA	PG	1.0
678	539.0	Alondes Williams	23.0	BRK	SG	1.0
679	Unknown	League Average	Unknown	Unknown	Unknown	Unknown

	GS	MP	FG	FGA	...	DRB	TRB	AST \
0	66.0	34.6	11.0	20.1	...	8.4	10.2	4.2
1	66.0	36.2	10.9	22.0	...	7.8	8.6	8.0
2	58.0	36.3	9.6	20.7	...	4.0	4.8	7.3
3	68.0	35.5	10.4	20.3	...	4.0	4.8	5.5
4	63.0	32.1	11.2	20.3	...	9.6	11.8	5.7
..
675	0.0	3.0	0.0	1.0	...	0.0	0.0	0.0
676	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0
677	0.0	5.0	0.0	3.0	...	1.0	2.0	1.0
678	0.0	5.0	0.0	0.0	...	1.0	1.0	0.0
679	Unknown	Unknown	Unknown	Unknown	...	Unknown	Unknown	Unknown

	STL	BLK	TOV	PF	PTS	Awards \
0	1.0	1.7	3.4	3.1	33.1	MVP-1DPOY-9CPOY-5ASNBA1
1	1.4	0.5	3.6	2.5	32.4	MVP-8CPOY-8ASNBA1
2	0.9	0.3	3.3	1.9	32.2	CPOY-10ASNBA3
3	1.6	1.0	2.8	2.8	31.4	MVP-5CPOY-7ASNBA1
4	0.8	0.8	3.9	3.1	31.1	MVP-3DPOY-6ASNBA1
..
675	0.0	0.0	0.0	0.0	0.0	Unknown
676	0.0	0.0	0.0	0.0	0.0	Unknown
677	0.0	0.0	0.0	0.0	0.0	Unknown
678	0.0	0.0	2.0	1.0	0.0	Unknown
679	Unknown	Unknown	Unknown	Unknown	Unknown	Unknown

```
      Player-additional
0      embiijo01
1      doncilu01
2      lillada01
3      gilgesh01
4      antetgi01
..      ...
675     burtode02
676     fostemi02
677     jacksfr01
678     willial06
679     -9999
```

```
[680 rows x 32 columns]
All data is valid.
```

```
[ ]:
```